

Automated Transcription of Jazz Soli

Back to roots of automated music transcription - proposal of M2 internship

supervisors: [Florent Jacquemard](#), [Philippe Rigaux](#), [Francesco Foscari](#)

laboratory: équipe [Vertigo](#), [Cedric](#), [CNAM](#), 2 rue Conté 75003 Paris.

summary: The purpose of Automated Music Transcription (**AMT**, [1]) is to convert performed music into music notation (*i.e.* music scores). This ancient problem of computer music has its origin in the desire to preserve, in a written form, improvised music or music of oral tradition, and to ease exchanges with musicians, students or scholars (musicologists). Current applications include music writing and edition, education, production (content visualisation), search (indexing of databases based on symbolic representations) and musicology. AMT involves many subtasks that became research problems by themselves in the field of Music Information Retrieval (MIR); It is still considered a challenging and open problem in the literature.

The goal of this internship is to develop methods for the automated transcription of jazz solo improvisations, starting from symbolic representations of note pitches over time (called **piano-roll** representations [2]). The latter are often considered as **intermediate representations**, extracted from audio recordings by signal processing, in full-scale (end-to-end) transcription scenarios see [1, Fig.1].

We propose an approach using the two following elements to leverage this problem.

- The **dataset** [3] from the Weimar Jazz Database (Jazzomat project), made of hundreds of jazz soli in different formats, in particular precise piano roll representations of the notes and timings in actual recorded performances, and score transcriptions. The datasets of the project Dig That Lick [4] (follow-up of Jazzomat) could also be considered.
- A **framework** called `qparse` [5] for the transcription of piano-rolls into music scores that solves jointly, in one pass, the two subtasks of rhythm quantization and music score structuring. It is based, throughout the transcription process, on *a priori* music language models (**MLM**), given in the form of Weighted Tree Automata (**WTA**). Such models describe the notations expected with corresponding ranking values; They can be trained on a corpus of music scores [6]. Transcription is then performed by algorithms for quantitative parsing and on-the-fly automata construction.

Note that the scores of the dataset [3] have been typesetted manually (from audio recordings) during the Jazzomat Research Project, with the help of `Sonic Visualiser` and the `MeloSpySuite`. The project was indeed not successful on performing totally automatic transcription, which is considered difficult for this case study, due in particular to the intrinsic rhythmic complexity encountered in jazz soli, with unusual tuplets, syncopes, few pattern regularities *etc.*

The work will comprise the following tasks:

1. **training** of MLMs from the chosen dataset,
2. **adaptation** of the above framework [1] and its current MLMs to the problem of transcribing of piano-rolls of jazz solo improvisations into music scores,

3. **evaluation** of the transcription performances on datasets.

The importance of these three complementary activities during the internship will fit the affinities of the student. One possibility for a theoretically inclined candidate could be to focus on 1. and study appropriate techniques for learning WTA. That includes in particular methods for training the weights of WTA, when the state transitions are either fixed or can be slightly adapted, *e.g.* by state fusing or splitting [6]. Or algebraic methods for learning WTA (weights and state transitions), or *rational tree series*, from sample sets [7,8]. Another option is to focus on 2. and *n*-best enumeration algorithms for *rational tree series* [10].

Alternatively, she or he could favour experimental activities for the development and evaluation of a robust transcription procedure for Jazz soli.

A further problem of interest is the study of the relationship between acoustic models and language models as above in the context of the longer term problem of **end-2-end** transcription of jazz soli (from audio recordings into music scores).

prerequisites: It is preferable (but not mandatory) for the candidate to have backgrounds on formal language theory and on music notation and other symbolic music representations.

references

- [1] Automatic Music Transcription: An Overview. Emmanouil Benetos, Simon Dixon, Zhiyao Duan, Sebastian Ewert. IEEE Signal Processing Magazine 26(1), 2019.
- [2] Fundamentals of Music Processing. Meinard Müller. Springer 2015. *see also* https://www.audio-labs-erlangen.de/resources/MIR/FMP/C1/C1S2_PianoRoll.html
- [3] Weimar Jazz Dataset <https://jazzomat.hfm-weimar.de>
- [4] DTL1000 database (Dig That Lick project) <http://dig-that-lick.eecs.qmul.ac.uk>
- [5] A Parse-based Framework for Coupled Rhythm Quantization and Score Structuring. Francesco Foscarin, Florent Jacquemard, Philippe Rigaux, Masahiko Sakai. Conference on Mathematics and Computation in Music (MCM), 2019. *see also* <https://qparse.gitlabpages.inria.fr>
- [6] Modeling and Learning Rhythm Structure. Francesco Foscarin, Florent Jacquemard, Philippe Rigaux. Sound and Music Computing Conference (SMC), 2019.
- [7] A Formal View on Training of Weighted Tree Automata by Likelihood-Driven State Splitting and Merging. Toni Dietze. PhD univ. Dresden, 2018.
- [8] Spectral Learning of Weighted Automata. Borja Balle, Xavier Carreras, Franco M Luque, and Ariadna Quattoni. Machine learning, 96(1-2):33–63, 2014.
- [9] Complexity of Equivalence and Learning for Multiplicity Tree Automata. Ines Marušić, James Worrell. Journal of Machine Learning Research, 2015.
- [10] Efficient Enumeration of Weighted Tree Languages over the Tropical Semiring. Johanna Björklund, Frank Drewes, Niklas Zechner. Journal of Computer and System Sciences, vol. 104, 2019.